

ANNA STRZEBOŃSKA

Dowody w zarządzaniu miastem. Czy w decyzjach publicznych jest miejsce na naukę?¹

Generalną tezę artykułu jest twierdzenie, że decyzje o charakterze publicznym podejmowane w polityce miejskiej powinny być oparte na najlepszych dostępnych dowodach. Czy jednak są te „dowody”?

Ciężar dyskusji na temat zasadności oceny miejskich interwencji publicznych można przenieść na grunt rozważań na temat tego, co dokładnie rozumie się przez pojęcie „dowód” (*evidence*). Sięgając do definicji *Słownika języka polskiego*² za dowód uznaje się „okoliczność lub rzecz dowodzącą czegoś, świadczącą o czymś”. Innymi słowy, dowody mogą być wykorzystane na poparcie wniosku, nie są jednak z nim tożsame. Dowody nie mają mocy podejmowania decyzji. Decyzję podejmuje ludzie w danym kontekście, np. twórcy miejskich polityk publicznych. Takie rozumienie dowodów ma wiele konsekwencji.

Sposób rozumienia dowodów zależy od roli zawodowej. Dla badacza dowody oznaczają zagregowane dane (zarówno jakościowe, jak i ilościowe) uzyskane w procesie badawczym opartym na syntezie, których prawdziwość została potwierdzona przez przeprowadzoną procedurę oceniającą wiarygodność naukową uzyskanego materiału empirycznego. Wybór wspomnianej procedury determinowany jest podjętymi rozstrzygnięciami o charakterze metodologicznym,

¹ Punktem wyjścia tego tekstu była moja rozprawa doktorska pt. *Metaanaliza w ewaluacji programów społecznych*, napisana pod kierunkiem prof. dr. hab. Jarosława Górniaka, sfinansowana w ramach grantu badawczego promotorskiego Komitetu Badań Naukowych (numer projektu N N116 284438) ze środków Ministerstwa Nauki i Szkolnictwa Wyższego.

² Na podstawie informacji umieszczonych w elektronicznej wersji *Słownika języka polskiego* dostępnego na stronie www.sjp.pwn.pl.

dotyczącymi np. zastosowanego schematu badawczego, ale powinien posiadać cechy wiedzy naukowej.

Twórcy miejskich polityk publicznych nie myślą o dowodach w taki sam sposób jak badacze. Poszukują oni informacji, które pozwolą im odpowiedzieć na następujące pytania³ (Davies 2006: 97–104):

- Czy miejska polityka publiczna X osiągnęła swoje cele? (pytania zorientowane na cele);
- Czy miejska polityka publiczna X okazała się skuteczniejsza od miejskiej polityki publicznej Y pod kątem osiągniętych efektów? (pytania porównujące skuteczność);
- Jaki jest najskuteczniejszy sposób wdrażania miejskiej polityki publicznej X? (pytania zorientowane na wdrażanie);
- Jakie są bariery w procesie wdrażania miejskiej polityki publicznej X oraz osiągania postawionych przed nią celów? (pytania zorientowane na wdrażanie);
- Jakie są koszty i korzyści miejskiej polityki publicznej X w porównaniu z miejską polityką publiczną Y? (pytania porównawcze o charakterze ekonomicznym);
- Czy miejska polityka publiczna X jest bardziej efektywna kosztowo (opłacalna) w osiąganiu wyznaczonych celów niż miejska polityka publiczna Y? (pytania porównawcze o charakterze ekonomicznym);
- Jaki jest wpływ miejskiej polityki publicznej X (Y/Z) na środowisko naturalne, biznes, administrację, społeczną równość itd.? (pytania o wpływ);
- Jakie inwestycje/działania muszą zostać poniesione, by utrzymać efekty (zapewnić trwałość) miejskiej polityki publicznej X? (pytania o charakterze ekonomicznym i strategicznym);
- Czy należy wdrożyć miejską politykę publiczną X, jeżeli oznacza to zaniechanie wdrożenia polityki publicznej Y? (pytania o charakterze etycznym i strategicznym);
- Jak nasze miasto wypada na tle podobnych miast pod kątem zmian społecznych i gospodarczych? (pytania o charakterze audytu strategicznego, pozwalające na dokonywanie porównań pomiędzy miastami).

Autorzy kanadyjskiego badania na temat postrzegania dowodów w służbie zdrowia – Jonathan Lomas i in. – ustalili, że poza środowiskiem naukowym dominuje nienaukowe podejście do dowodów, rozumianych jako „wszystko, co

³ Prezentowane pytania zostały opracowane na gruncie brytyjskim, na podstawie współpracy rządu (a dokładnie The Government Chief Social Researcher's Office in the Cabinet Office) z badaczami zrzeszonymi w ramach ESRC Evidence Network.

ustanawia fakt lub daje podstawę do wiary w coś” (2005: 3). Naukowcy są bardziej restrykcyjni i ograniczają pojęcie „dowodów” do informacji uzyskanych za pośrednictwem procesów i procedur badawczych uznanych za naukowe. Na podstawie tego odkrycia opracowali typologię dowodów utworzoną pod kątem sposobów ich uzyskania (Lomas i in. 2005: 3–4):

1. Dowody wolne od kontekstu (*context-free evidence*) oznaczają informacje, które mają charakter uniwersalny i zostały uzyskane w wiarygodnym procesie badawczym (wysokiej jakości). Dostarczają wiedzy ogólnej, np. na temat zachowania ludzkiego organizmu pod wpływem konkretnego bodźca, i są niezmiennie w każdych warunkach (np. społecznych, kulturowych) pomiaru. Najczęściej spotykane są w naukach medycznych do oceny skuteczności interwencji oraz w badaniach biomedycznych (np. farmaceutycznych, kosmetycznych, psychologicznych).
2. Dowody wrażliwe na kontekst (*context-sensitive evidence*) oznaczają informacje, które nie mają charakteru uniwersalnego, ale zostały uzyskane w wiarygodnym procesie badawczym (wysokiej jakości). Ich cechą charakterystyczną jest to, że nie są odporne na pojawiające się różnice między warunkami (np. społecznymi, kulturowymi) przeprowadzenia pomiaru. Ten typ dowodów najczęściej spotyka się w naukach społecznych, a także w badaniach hybrydowych, np. łączących odkrycia biomedyków z naukami społecznymi lub równocześnie wykorzystujących metody jakościowe i ilościowe (triangulacja metod) bądź syntetyzujących ustalenie badawcze w procesie przeglądu systematycznego.
3. Dowody potoczne (*colloquial evidence*) oznaczają informacje, które mogą mieć zarówno charakter uniwersalny, jak i – zdecydowanie najczęściej – nieuniwersalny. Charakteryzują się tym, że nie zostały wytworzone w procesie badawczym, a ich źródłem są opinie ekspertów (które mogą być subiektywne) i/lub indywidualne doświadczenia, które mogą mieć jedyny w swoim rodzaju charakter.

Z przedstawionej typologii tylko dwa pierwsze rodzaje dowodów – wolne od kontekstu i wrażliwe na kontekst – identyfikowane są jako naukowe, gdyż ich wiarygodność jest sprawdzona lub sprawdzalna pod względem metodologicznym. To, co je odróżnia, to zakres, w jakim mogą być uogólnione. Najbardziej problematyczne – oczywiście z punktu widzenia metodologa – są tzw. dowody potoczne. Ich zastosowanie determinuje jedynie odpowiedniość kontekstowa (trafność). Nie są sprawdzone pod kątem wiarygodności. W literaturze (Carr 1992: 241–251) pojawiają się głosy, że informacje tego typu powinny mieć charakter uzupełniający wobec pozostałych typów dowodów, szczególnie w sytuacji, kiedy doniesienia badawcze są niejednoznaczne bądź zostały uzyskane w procesie badawczym niskiej jakości. Ideę tę ilustruje schemat 1.



SCHEMAT 1. Wzajemna zależność dowodów potocznych i tych o naukowym charakterze

Źródło: opracowanie na podstawie Lomas i in. 2005: 3.

Niestety, w naszym kraju najczęściej wykorzystywane przez twórców miejskich polityk publicznych są dowody o charakterze potocznym. Niewątpliwie na taki stan rzeczy wpływ ma marginalny pomiar związków przyczynowych w ewaluacji, który – mimo najlepszych chęci i zastosowanej metodologii pomiaru – cechuje duże zróżnicowanie w zastosowanych schemacie badawczym, które uniemożliwia zastosowanie integracji danych.

W jaki sposób wytworzyć wiarygodne dowody w ocenie miejskich polityk publicznych?

Jeżeli proces miejskich interwencji zaplanowanych w ramach polityk publicznych jest rozłożony w czasie i „może być zdefiniowany jako systematyczna seria działań skierowanych na osiągnięcie założonych celów” (Pincus i Minahan 1973: 85), to ocena wpływu interwencji (*impact evaluation*) następuje po jej całkowitym wdrożeniu, czyli na etapie ewaluacji *ex-post*, a koncentruje się na ocenie „pozytywnych i negatywnych, pierwotnych i wtórnych długoterminowych efektów wyprodukowanych przez tę interwencję w sposób bezpośredni i pośredni, zamierzony i niezamierzony” (OECD–DAC 2002: 22–23). Pomiar

„skutków ubocznych” wymusza przyjęcie innych kryteriów niż skuteczność interwencji, gdyż z definicji ocena skuteczności ogranicza się tylko do weryfikacji zaplanowanych efektów poprzez kontrolę stopnia realizacji założonych przed jej wdrożeniem wartości wskaźników. Z tego powodu w 2004 roku pojawiło się nowe kryterium oceny, tzw. użyteczność sprawdzająca odniesienie wszystkich efektów interwencji publicznych do potrzeb lub problemów wymagających rozwiązania (DG Budget Evaluation Unit 2004). Rodzaje efektów wraz z ich definicjami przedstawia schemat 2.

Ocena wpływu miejskich interwencji (*impact evaluation*) o charakterze społecznym pod kątem użyteczności powinna przebiegać na dwóch poziomach wyróżnionych z punktu widzenia jednostki analizy.

Poziom I – rozumiany jako pomiar i analiza efektów miejskiej interwencji w wymiarze indywidualnym, czyli w stosunku do jednostek, które skorzystały z oferowanych przez nią form wsparcia wraz z wyjaśnieniem przyczyn ich występowania. Na tym poziomie wymagane jest także wyjaśnienie przyczyn zaobserwowanych wartości efektów interwencji. Jednostką analizy są podmioty zdefiniowane jako grupa docelowa interwencji (osoby, grupy osób, organizacje itp.), a wartość efektów odnosi się do wyzwań bądź celów na poziomie mikro (*objectives*).

Poziom II – czyli synteza i analiza efektów miejskich interwencji oszacowanych na poziomie I (które w postaci zagregowanej nazywane są dowodami) wpływa na sytuację ekonomiczną, społeczną i gospodarczą całego miasta lub innego obszaru terytorialnego, na którym została wdrożona interwencja, np. obszaru metropolitalnego. Na tym poziomie konieczne jest również zagregowanie danych wyjaśniających, zaobserwowane na poziomie I wartości efektów interwencji. Jednostką analizy są dane wtórne, czyli doniesienia (wyniki, wnioski i rekomendacje) ewaluacji zrealizowanych na poziomie I. Dowody odnoszą się do wyzwań bądź celów na poziomie makro (*goals*).

Innymi słowy, dowody tworzy się poprzez syntezę wniosków z różnego typu badań, których jakość warunkuje bezpośrednio jakość dowodów. To, czy mamy do czynienia z naukowymi, czy potocznymi dowodami, zależy od rozstrzygnięć, jakie badacze bądź zamawiający poczynili na etapie projektowania badań ewaluacyjnych mających udzielić odpowiedzi na pytania: Jaka jest wartość efektów miejskich interwencji?; Jak rozstrzygnąć, czy interwencja publiczna przyniosła sukces czy porażkę?; Jak duży – pozytywny i/lub negatywny – był ten efekt?

Mimo pozornej prostoty pytania nie jest łatwo udzielić na nie odpowiedzi. Jak argumentuje Earl Babbie:

potencjalnie jednym z najbardziej uciążliwych aspektów badań ewaluacyjnych jest ustalenie, czy badany program odniósł sukces, czy poniósł porażkę (...) Jak można się spodziewać, nie ma ostatecznych odpowiedzi na tego typu pytania (2003: 375).

Prekursorem podejścia do ewaluacji patrzącej szerzej niż osiągnięte założone cele programu był Ralph W. Tyler (1942: 492–501). Jego koncepcja została

rozwinięta w latach 70. ubiegłego stulecia, m.in. przez Michaela Scrivena (1972: 1–7), który argumentował, że w ewaluacji programów publicznych nie należy koncentrować uwagi tylko na założonych do realizacji celach. Równie ważne są wszystkie inne – nawet nieprzewidziane – efekty programu, na których pomiarze powinna skupić się uwaga ewaluatorów⁴. Opracował on podejście teoretyczne zwane „ewaluacją wolną od celów”.

Tym, co może zrealizować prawidłowo przeprowadzona ewaluacja *ex-post*, to ustalenie przez spójną metodykę opartą na wysokim standardzie metodologicznym, pozwalającą na obiektywną, trafną, rzetelną i standaryzowaną ocenę analizowanych działań interwencji związków przyczynowych⁵ między obserwowanymi wynikami a podjętą interwencją, w których kluczową kwestią jest pomiar efektu netto⁶. Obliczenie efektu netto umożliwia porównanie efektów programu z sytuacją, w której dane działanie nie miałoby miejsca⁷. Taką analizę efektu netto w ramach ewaluacji rekomendował już w 1997 roku podręcznik wydany przez Komisję Europejską:

Gdy mówimy, że dane efekty zostały wyprodukowane lub spowodowane przez program, oznacza to, że gdyby program nie miał miejsca lub przeprowadzano by go w innej formie lub stopniu, te efekty nie wystąpiłyby lub nie wystąpiłyby w tym samym stopniu. Oznacza to, że ważne jest, by mieć precyzyjny obraz tego, co stałoby się bez programu (Dyrekcja Generalna XIX – Budżet 1997: 39 za: Górniak i Keler 2008: 118).

Badanie związków przyczynowych składa się z dwóch etapów, które polegają na (Heckman i Smith 1996: 38; Górniak 2009: 192):

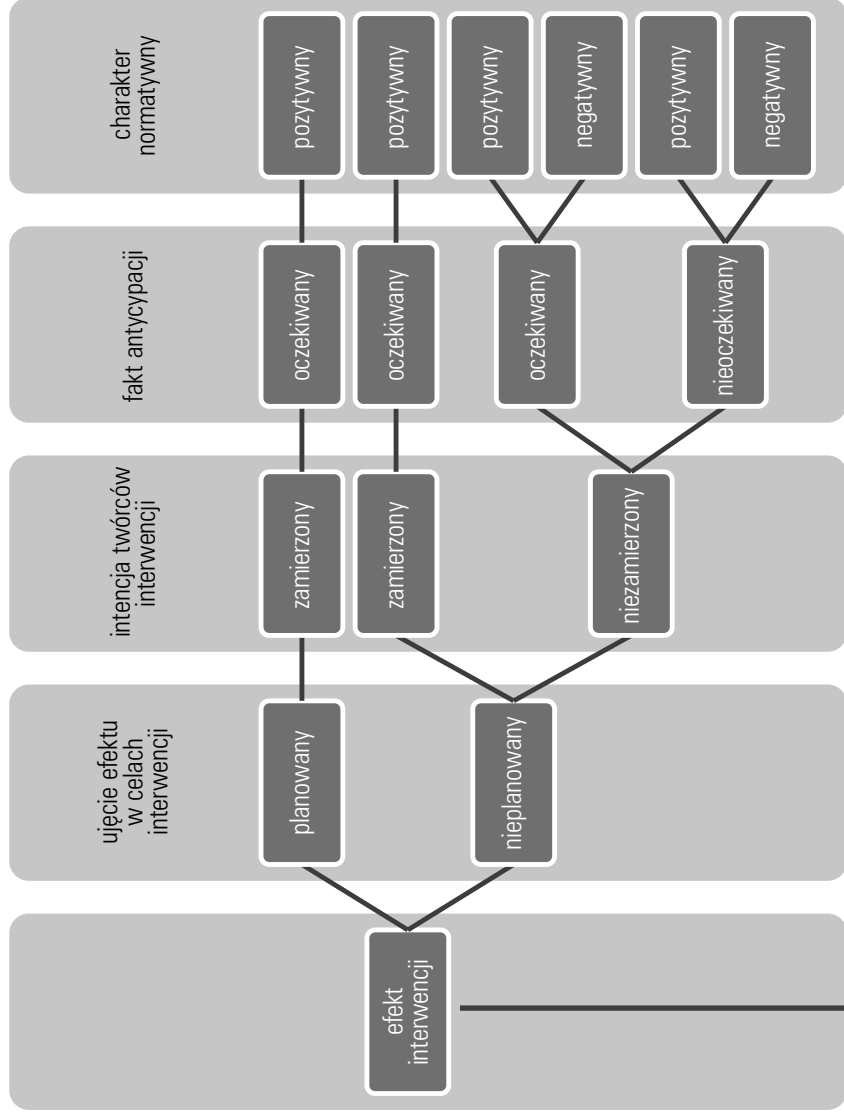
- zmierzeniu **efektu brutto** miejskiej interwencji jako ogólnej zmiany zdefiniowanej w jej celach (za pomocą ustalonych wskaźników efektywności) poprzez udzielenie odpowiedzi na pytanie, jakie są efekty interwencji w stosunku do uczestniczących w niej jednostek,
- oddzieleniu zmiany, która wystąpiła niezależnie od wdrożonej miejskiej interwencji, od zmiany wywołanej tylko przez tę interwencję (określenie **efektu netto**) poprzez udzielenie odpowiedzi na pytanie, jak w wyniku realizowania zaplanowanych w interwencji działań zmienia się pozycja jednostek objętych wsparciem w relacji do pozycji, jaką mieliby, gdyby nie uczestniczyli w programie.

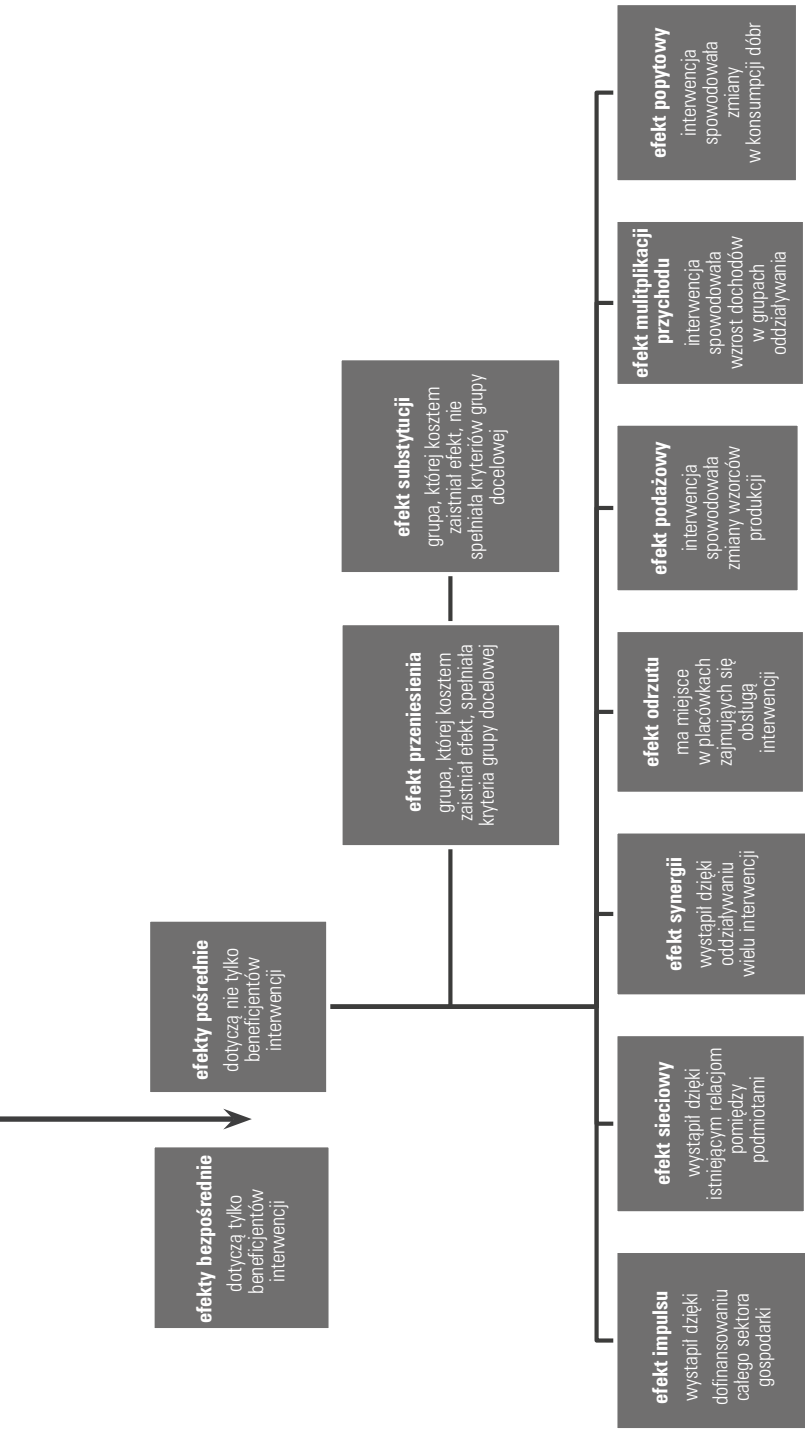
⁴ Scriven argumentował, że pomiaru rzeczywistych efektów programu nie powinna zakłócić jego retoryka. Z tego powodu zadaniem ewaluatorów jest unikać czytania broszur czy też bardziej rozbudowanych opisów ocenianego programu.

⁵ Związek przyczynowy ustala się w ten sposób, że interwencję traktuje się jako bodziec, a efekt jako jego skutek.

⁶ Metodykę wyliczania efektu netto jako miary przyczynowego efektu interwencji, jak i problemy związane z obliczaniem tejże miary, można znaleźć w pracy Jarosława Górniaka (2009: 194–197).

⁷ Tak zwana „sytuacja kontrfaktyczna”, zob. Dyrekcja Generalna XIX – Budżet 1997: 39.





SCHEMAT 2. Pięciowymiarowa klasyfikacja efektów interwencji

Źródło: opracowanie na podstawie Górniak i in. 2010: 18–21.

Koncept pomiaru efektu netto interwencji został zapożyczony z nauk medycznych i opiera się na metodologii badań eksperymentalnych i quasi-eksperymentalnych⁸. Wyliczenia różnicy w wartościach ustalonych miar (wskaźników) pomiędzy grupą jednostek, które zostały poddane interwencji (zwaną dalej grupą interwencyjną) i grupą jednostek pozostających poza jej oddziaływaniem (zwaną dalej grupą kontrolą) mają być „wyrażone w liczbach i bazować na **wiarygodnych** danych empirycznych” (European Cohesion Fund i European Regional Development Fund 2011: 5).

Dodatkowo, korzystając z metodologii opracowanej w ramach badania pt. *Ewaluacja oparta na badaniu użyteczności – analiza w kontekście realizacji zasady good governance w wybranych programach operacyjnych* (Górnjak i in. 2010), warto zaznaczyć, że efekt netto interwencji może przyjmować dwie wartości – pozytywną i negatywną. Negatywny efekt netto (wyrażony w wartościach ujemnych) oznacza antyżyteczność interwencji społecznej, natomiast pozytywny efekt netto (wyrażony w wartościach dodatnich) świadczy o użyteczności podejmowanych działań, a zakres ich sukcesu odzwierciedlają wartości efektu.

W celu najlepszego pomiaru wpływu interwencji publicznych warto wykorzystać podejście zaproponowane w 2002 roku przez Davida Storeya o nazwie „sześć kroków do nieba”, opracowane na podstawie doświadczeń brytyjskich w ewaluacji i monitoringu (2002: 181–202). Zakłada ono kompatybilne wykorzystanie danych z systemu monitoringu wdrażanych działań (np. rejestry miejskie) i danych pozyskanych w badaniu ewaluacyjnym. Koncepcja ma charakter kumulacyjny, co oznacza, że każdy kolejny krok przybliża badacza do ideału reprezentowanego przez szósty poziom (tabela 1).

Piąty i szósty krok wymagają symulowania sytuacji kontrfaktycznej, czyli wyliczeń wartości efektów netto. Osiągnięcie „nieba” nie neguje konieczności monitoringu wdrażania zaplanowanych działań, a wręcz uzasadnia potrzebę jego istnienia. Monitoring dostarcza ważnych danych na temat zgodności zaplanowanych działań, czyli „środków” do osiągnięcia założonych celów. Dodatkowo, ocena tej zgodności ważona jest przez pryzmat opinii odbiorców interwencji, którzy oceniają jej trafność (a raczej przydatność praktyczną), czasochłonność i poziom skomplikowania procedur wdrażania, a także wyrażają swoje subiektywne zadowolenie z uczestnictwa w konkretnej formie wsparcia. Wszystkie te informacje są ważne, jednak nie powinny być dominujące na etapie ewaluacji *ex-post*. Ocena procesu wdrażania ma sens wówczas, gdy istnieje

⁸ Interesujący jest fakt, że we wspomnianym na początku pracy podręczniku do metodologii badań społecznych autorstwa E. Babbiego, wydany w 2003 roku, znajduje się rozdział poświęcony badaniom ewaluacyjnym prowadzonym metodami eksperymentalnymi i quasi-eksperymentalnymi. Zatem wiedza naukowa na temat sposobów prowadzenia badań wpływu interwencji społecznych istniała na rynku krajowym jeszcze przed naszym członkostwem w Unii Europejskiej. Niestety nikt z niej nie skorzystał.

szansa na zastosowanie proponowanych w ewaluacji ulepszeń, czyli w trakcie prowadzenia interwencji (ewaluacja *on-going*). Natomiast kwestia pomiaru satysfakcji odbiorców jest kontrowersyjna z punktu widzenia metodologicznej wewnętrznej trafności badania, bo może generować obciążenia wyników pomiaru w sytuacji, kiedy respondenci zdają sobie sprawę, że instytucją zlecającą badanie jest instytucja przyznająca środki. Zatem by nie zamykać sobie drogi do potencjalnego przyszłego wsparcia, odbiorcy „na wszelki wypadek” deklarują wysoką satysfakcję i/lub przydatność formy działania interwencji, w której uczestniczyli.

TABELA 1. Koncepcja „sześciu kroków do nieba” Davida Storeya

DANE Z MONITORINGU	<p>1 krok: Ocena schematu miejskiej interwencji publicznej ze szczególnym uwzględnieniem form wsparcia, ich adresatów i kryteriów selekcji uczestników.</p> <p>2 krok: Opinie odbiorców wsparcia na temat trafności oferowanych działań (ocena schematu interwencji) oraz procedur ich wdrożenia (pod kątem takich kryteriów, jak czasochłonność, skomplikowanie, satysfakcja).</p> <p>3 krok: Opinie odbiorców wsparcia na temat efektów, jakie przyniósł im udział w konkretnej formie miejskiej interwencji (deklarowany efekt netto).</p>
DANE Z EWALUACJI	<p>4 krok: Porównanie uzyskanych efektów u odbiorców wsparcia z sytuacją przeciętego podmiotu (np. na podstawie ogólnokrajowych statystyk).</p> <p>5 krok: Porównanie sytuacji odbiorców wsparcia (uzyskanych efektów) z sytuacją podobnych jednostek, które nie uczestniczyły w interwencji, czyli zastosowanie grup kontrolnych.</p> <p>6 krok: Porównanie sytuacji odbiorców wsparcia (uzyskanych efektów) z sytuacją takich samych jednostek (o takich samych charakterystykach), które nie uczestniczyły w interwencji z użyciem technik statystycznych kontrolujących obciążenie uzyskanych wyników na podstawie zastosowanego procesu selekcji (<i>selection bias</i>).</p>

Źródło: opracowanie na podstawie Storey 2002: 181–202.

Warto podkreślić, że pomiar efektów netto jest stabilniejszy pod kątem metodologicznej trafności pomiaru i co najważniejsze, pozwala wytworzyć dowody. Tego typu opinie pojawiły się już w 2004 roku w obszarze funduszy strukturalnych. W tym czasie Komisja Europejska zleciła dwójce holenderskich badaczy – Marco Mosselmanowi i Yvonne Prince – poszukiwanie odpowiedzi na pytanie dotyczące tego, która metoda ewaluacji jest najwiarygodniejsza w ocenie wpływu interwencji publicznych. Po dokonaniu przeglądu literatury

naukowej z dziedziny ewaluacji oraz raportów z tego typu badań, badacze opracowali kwestionariusz ankiety zawierający wiele pytań na temat tego, jakie cechy powinno posiadać wiarygodne badanie oceniające skuteczność publicznego wsparcia. Opracowane narzędzie wykorzystali do poznania opinii 19 państw członkowskich UE oraz zidentyfikowania podejść badawczych, które są wykorzystywane w ewaluacji polityk publicznych w danym kraju. Wyniki badania omówiono 25 maja 2004 roku w Brukseli, na spotkaniu przedstawicieli państw członkowskich Unii Europejskiej⁹, na którym ustalono, co należy robić w procesie pomiaru efektów interwencji publicznych (Mosselman i Prince 2004: 61–62):

- Unikać badań deklaracyjnych odbiorców wsparcia, mimo że są pożądane przez społeczeństwo¹⁰. Warto pamiętać, że osoba objęta interwencją nigdy nie wyrazi opinii, że otrzymana przez nią pomoc była zbędna¹¹.
- Korzyści z badań ewaluacyjnych powinny być nadrzędne nad kosztami i ograniczeniami (np. zasobów). Z tego powodu warto używać skomplikowanych, ale kompleksowych metod oceny (np. ewaluacji wolnej od celów).
- Zidentyfikować i opisać „straty uboczne”, czyli negatywne efekty (zaplanowane bądź też nie), tj. rozwój biurokracji, wzrost obciążeń administracyjnych wśród podmiotów wdrażających i otrzymujących wsparcie czy choćby zakłócenia w funkcjonowaniu „wolnego rynku”.
- Zakłócenia w funkcjonowaniu „wolnego rynku” powinny być mierzone z wykorzystaniem zjawiska przeniesienia (*phenomenon of displacement*)¹², a wynik analizy powinien znaleźć się w raporcie z ewaluacji.

⁹ Na spotkaniu byli obecni przedstawiciele następujących państw: Austrii, Belgii, Cypru, Danii, Estonii, Finlandii, Francji, Grecji, Hiszpanii, Irlandii, Łotwy, Litwy, Niderlandów, Niemiec, Norwegii, Polski, Portugalii, Słowacji, Słowenii, Szwecji, Węgier, Wielkiej Brytanii, Włoch. Natomiast do krajów, które w 2004 roku przeprowadzały ewaluację wpływu wszystkich interwencji publicznych należą: Irlandia, Niderlandy (Holandia), Słowacja – są to państwa o największym doświadczeniu w pomiarze skuteczności programów publicznych. Natomiast jedynie elementy ewaluacji wpływu części programów publicznych były podejmowane w: Austrii, Belgii, Republice Czeskiej, Danii, Finlandii, Niemczech, Norwegii, Portugalii, Szwecji, Węgrzech, Wielkiej Brytanii, Włoszech. Źródło danych: Mosselman i Prince 2004: 62–64.

¹⁰ A przynajmniej taki obraz kreowany jest w mediach, w których zdanie opinii publicznej na aktualnie ważne tematy wyrażane jest za pomocą sondaży opinii (wszelkiego rodzaju grafiki, np. wykresy słupkowe).

¹¹ Taką opinię potwierdzają wyniki praktycznie wszystkich badań ewaluacyjnych w Polsce, według których wszyscy odbiorcy wsparcia uważają, że otrzymana pomoc była potrzebna.

¹² Za efekt przeniesienia (przemieszczenia) w literaturze przedmioty uznaje się: „Efekt uzyskany na obszarze kwalifikującym się do udzielenia pomocy kosztem innego obszaru. Efekt przeniesienia może być zamierzony (np. przeniesienie placówek administracji publicznej ze stolicy na obszar gospodarczo »zacořany«) lub niezamierzony (np. 10% stworzonych dzięki programowi rozwoju regionalnego miejsc pracy przyczyniło się do zniknięcia miejsc pracy na innych kwalifikujących się obszarach, gdyż firma (dzięki programowi) przeniosła swoją

- Ewaluacja procesu wdrażania powinna mieć miejsce w trakcie wdrażania interwencji publicznej (tzw. ewaluacja *on-going*), aby mieć czas na wprowadzenie ulepszeń w planie realizowanego wsparcia.
- Należy uwzględnić kontekst wprowadzenia interwencji publicznej na poziomie makro, tj. sytuacji gospodarczej kraju, zadań strategicznych państwa czy jakości dostępnych danych, która warunkuje identyfikację najlepszych typów wsparcia.

W wyniku prowadzonych prac badawczych wyłoniły się wskazówki dla zlecających ewaluację wpływu interwencji publicznych. Powstały one na zasadzie prostej dychotomii działań – identyfikującej te, które należy wykonać, przeprowadzając badania ewaluacyjne i te, których bezwzględnie powinno się unikać. Ich zbiorcze zestawienie prezentuje tabela 2.

TABELA 2. Wskazówki dla zlecających ewaluację wpływu interwencji publicznych

Wykonuj w ewaluacji	Unikaj w ewaluacji
<ul style="list-style-type: none"> • Zaczynij myśleć nad sposobem ewaluacji wpływu jeszcze przed rozpoczęciem wdrażania interwencji. Przypilnuj, by planowane formy interwencji miały określone wymierne cele operacyjne (<i>objectives</i>) i wskaźniki. • Zleć ewaluację niezależnemu badaczowi. • Przypilnuj, by ewaluacja była zgodna z założonymi celami operacyjnymi, ale wskazane jest, by wykraczała poza nie. • W ewaluacji łącz doniesienia metod i technik badawczych o charakterze jakościowym i ilościowym (triangulacja). • Zidentyfikuj wszystkich interesariuszy ewaluacji, ich interesy oraz oczekiwania względem doniesień badania (np. rodzaj efektów i/lub dowodów, który ich przekona). • Wykorzystuj metodologię doboru próby. Dopasuj odpowiednią wielkość próby do badań, nie za dużą, ale też nie za małą. Nie przeprowadzaj badań na próbie wyczerpującej, bo to zwiększa koszty. 	<ul style="list-style-type: none"> • Uważaj ewaluację wpływu za „zło konieczne”. • Odłóż myślenie o ewaluacji (np. o sposobach pomiaru efektów) na czas po zakończeniu interwencji. • Oczekuj nieobciążonych wniosków z autoewaluacji lub sondaży opinii wśród odbiorców wsparcia. • Ogranicz ewaluację do prostego pomiaru osiągnięcia zakładanych wartości wskaźników (<i>targets</i>). • W ewaluacji wykorzystuj tylko jedno źródło danych. • Myśl, że badanie ewaluacyjne jest proste i tanie. • Myśl, że badanie ewaluacyjne jest zbyt skomplikowane i zbyt kosztowne. • Stosuj standardowe procedury do oceny unikalnych interwencji.

siedzibę z centrum na obrzeża miasta). W przypadku efektu niezamierzonego, aby uzyskać efekt netto interwencji, należy odjąć efekt przeniesienia od efektów brutto. Termin »efekt przeniesienia« używany jest także czasem w odniesieniu do efektów wpływu interwencji na beneficjenta będącego na utrzymaniu innego beneficjenta na tym samym obszarze” (Portal Funduszy Europejskich 2012).

Wykonuj w ewaluacji	Unikaj w ewaluacji
<ul style="list-style-type: none"> • Używaj grup kontrolnych do pomiaru efektów interwencji (obliczaj efekty netto). • Uwzględnij obciążenie selekcyjne interwencji w pomiarze efektów netto. • Zintegruj badania ewaluacyjne z cyklem życia polityk publicznych. • Ucz się z poprzednich ewaluacji. 	<ul style="list-style-type: none"> • Zapomnij o pomiarze efektu występującego niezależnie (<i>deadweight</i>¹³), utworzeniu sytuacji kontrfaktycznej (<i>counterfactual</i>), efektu przeniesienia (<i>displacement</i>) oraz efektu dodatkowości (<i>additionality</i>¹⁴). • Zapomnij o pomiarze pozytywnych i negatywnych skutków ubocznych interwencji.

Źródło: opracowanie na podstawie Mosselman i Prince 2004: 40–41.

Jak wynika z tabeli 2, lekceważenie zalecanej metodologii pomiaru wpływu miejskich interwencji publicznych już na poziomie I uniemożliwia wiarygodne oszacowanie jej efektów, nie wspominając nawet o ich integracji. Przede wszystkim zagraża porównywalności efektów netto wytworzonych na poziomie I, gdyż zwiększa prawdopodobieństwo, że nie będzie możliwe porównywanie wartości osiągniętych efektów obliczonych w ramach pojedynczego badania ewaluacyjnego z wartościami efektów wyliczonych w innej ewaluacji – nie mówiąc już o porównaniach międzynarodowych, ze względu na zastosowany inny koncept lub sposób pomiaru.

Ponadto algorytm obliczania efektów interwencji może w ogóle nie bazować na pomiarze efektu netto, jeżeli nie zażyczy sobie tego zlecający badanie. Rzetelny pomiar efektu netto często jest sprzeczny z interesem instytucji

¹³ Efekt występujący niezależnie, tzw. *deadweight*, w nomenklaturze Komisji Europejskiej rozumiany jest jako „skutki, które wystąpiłyby nawet w sytuacji, gdyby program wydatków publicznych nie miał miejsca. *Deadweight* powstaje zazwyczaj jako rezultat nieadekwatnych mechanizmów realizacji programu (rozwiązania organizacyjne dostarczające określonym beneficjentom dobra i usługi finansowane przez program). Dzieje się tak zwłaszcza w sytuacji, gdy zrealizowane działania w programie nie przynoszą faktycznych korzyści zdefiniowanym grupom docelowym. W rezultacie inne osoby i grupy, które nie należą do grupy docelowej, stają się beneficjentami korzyści wygenerowanych przez program” (European Commission 2006).

¹⁴ Brouwer i in. definiuje dodatkowość (*additionality*) jako „sytuację, w której interwencja publiczna indukuje prywatne inwestycje, które nie byłyby podjęte bez otrzymanego wsparcia”. Ten efekt można mierzyć na trzy sposoby: 1) dotyczący wkładu (*input additionality*), który sprawdza, czy firmy prywatne wydają więcej swoich środków na zamierzone działania, bo były one dotowane, np. subsydiowane miejsca pracy, 2) dotyczący efektów (*output additionality*) – sprawdzenie, czy preferowana zmiana została dokonana, bo była dotowana (np. patenty, zatrudnienie), 3) dotyczący zmiany w zachowaniu (*behavioural additionality*) – czy formy wsparcia doprowadziły do trwałej zmiany zachowania. Obliczać go można, stosując modele ekonometryczne, ale najlepiej poprzez zastosowanie grup kontrolnych w pomiarze efektów interwencji (Brouwer i in. 2002).

zlecającej badanie. Zespół ekspertów¹⁵ w publikacji powstałej w 2012 roku na zlecenie Fundacji Gospodarki i Administracji Publicznej pt. *Kurs na innowacje. Jak wyprowadzić Polskę z rozwojowego dryfu?* wskazuje na patologiczny charakter samego modelu zamawiania badań ewaluacyjnych:

W praktyce instytucje zarządzające/wdrażające zamawiają raporty ewaluacyjne¹⁶ i zarazem je akceptują, co prowadzi do sytuacji, w której podmiot oceniany akceptuje raport dotyczący oceny efektów jego działalności, wpływając tym samym na jego ostateczny kształt (jeśli wziąć pod uwagę fakt, iż brak akceptacji wiąże się z możliwością obniżenia/wstrzymania wypłaty wynagrodzenia). Rynek badań ewaluacyjnych został zdominowany przez firmy konsultingowe (Geodecki i in. 2012: 62).

Takie zdominowanie rynku ewaluacji przez firmy konsultingowe – opierające się na prostym, nienaukowym podejściu do metodologii – przyczynia się do złamania zasady obiektywności badania i uniemożliwia tworzenie

wiarygodnej bazy do budowania szerszej wiedzy o efektach funduszy strukturalnych, a co za tym idzie – mechanizmach pożądanej zmiany społeczno-gospodarczej. Ewaluacja okazuje się prostym, biurokratycznym ćwiczeniem z opisu wydawania funduszy (Olejniczak 2009: 49).

Zdaniem wspomnianego wyżej zespołu ekspertów (Geodecki i in. 2012), model zamawiania badań ewaluacyjnych jest tylko jednym z przejawów słabości rozwojowej naszego państwa. Zdiagnozowane syndromy „choroby” wraz z ich cechami, a także konsekwencje praktyczne takiego stanu rzeczy oraz postulaty kierunku zmian, które pozwolą przezwyciężyć te słabości, prezentowane są w zbiorczym opracowaniu w tabeli 3. Z punktu widzenia ewaluacji wpływu interwencji społecznych najbardziej problematyczne są „słabości w sferze analizy, programowania i implementacji oraz oceny polityk publicznych” powiązane z „nikłymi kwalifikacjami klasy rządzącej do wytyczania celów rozwojowych i niską determinacją w ich realizacji”, co z kolei prowadzi do „niskiego stopnia powiązania wyników ewaluacji z procesami decyzyjnymi”. Brak przezwyciężania tych słabości utrzyma niewydolności systemu ewaluacji polityk publicznych i zablokuje rozwój metodologii badań w tej dziedzinie. Jeżeli twórcom miejskich

¹⁵ Do grona wspomnianych ekspertów należą: dr Tomasz Geodecki (UEK), prof. dr hab. Grzegorz Gorzelak (UW), dr hab. Jarosław Górniak, prof. UJ, prof. dr hab. Jerzy Hausner (UEK), dr hab. Stanisław Mazur (UEK), prof. dr hab. Jacek Szlachta (SGH) oraz prof. dr hab. Janusz Zaleski (IMGW/Politechnika Wrocławska).

¹⁶ Regulacja zgodna z rozdziałem 2. pkt 1 lit. B, „za realizację ewaluacji na poziomie programów operacyjnych odpowiedzialne są Instytucje Zarządzające”, istnieje w prawodawstwie w obowiązujących tzw. Wytycznych nr 6 w zakresie ewaluacji programów operacyjnych na lata 2007–2013 (Ministerstwo Rozwoju Regionalnego 2007).

interwencji nie zależy na pomocy w formułowaniu takich polityk publicznych, które mają szansę odnieść sukces (poprzez wiarygodne doniesienia badawcze), to jaki jest sens wykonywania badań ewaluacyjnych, które z definicji należą do badań stosowanych, czyli takich, których doniesienia wykorzystywane są w praktyce?

Zastanawiając się nad odpowiedzią na zadane pytanie, można odnaleźć cztery główne powody tego, że decydenci nie wykorzystują doniesień badawczych (w tym wniosków z badań ewaluacyjnych oceniających efekty wdrożonych interwencji) do planowania zakresu i kierunku przyszłych działań publicznych. Należą do nich:

- **niewiedza** na temat tego, jakie informacje są dostępne lub/i jakiego rodzaju ju działania powinni podjąć, by rozwiązać konkretny problem społeczny,
- **niezrozumienie** dostarczonych informacji, a dokładnie ich ważności i znaczenia w kontekście problemu społecznego do rozwiązania,
- **nieprzydatność** związana z postrzeganiem dostarczonych informacji jako bezużytecznych bądź niekorzystnych dla proponowanego rozwiązania problemu społecznego,
- **niezgoda** związana z postrzeganiem dostarczonych informacji jako fałszywych (nieodzwierciedlających rzeczywistości) i/lub błędnych (Bennett i Jessani 2001: 22).

Jeżeli nie uporamy się z ograniczeniami, to odpowiedzi na temat efektów interwencji będą cechowały się niską wiarygodnością metodologiczną i nie powinny (w myśl polityki opartej na dowodach) stanowić materiału dowodowego dla twórców polityk publicznych.

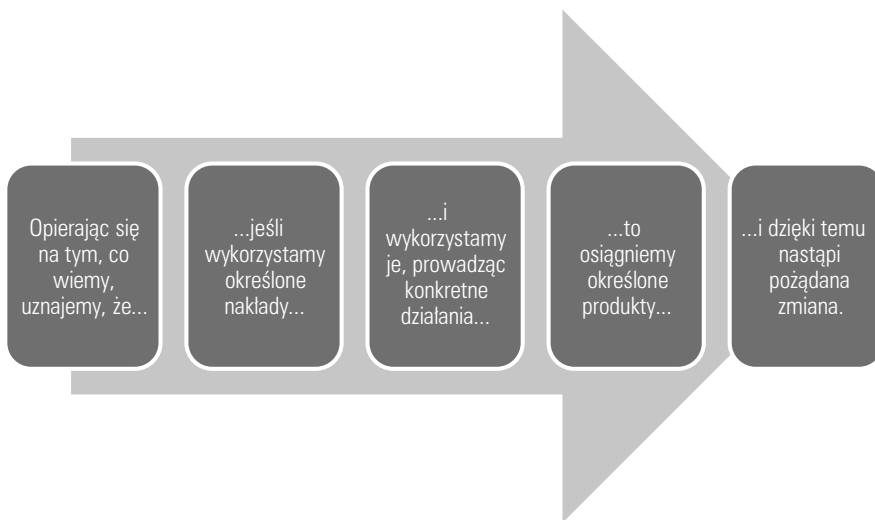
Ustalenie wielkości efektu miejskiej interwencji jest „połową drogi” do sukcesu zarządzania miastem opartego na dowodach. Kolejne kluczowe etapy stanowi udzielenie odpowiedzi na pytania: Dlaczego miejska interwencja osiąga zamierzone i niezamierzone efekty?; W jaki sposób działa?; W jaki sposób cechy interwencji wpłynęły na osiągnięte wartości efektów? U podstaw tych pytań leży założenie, że wartość efektów jest zależna od właściwości wewnętrznych (np. sposobów wdrożenia interwencji, czyli ustalenie, czy wszystko funkcjonowało zgodnie z planem) oraz zewnętrznych (np. kontekstu wprowadzenia interwencji).

Rekomendowane jest wykorzystanie teorii programu, która opiera się na powiązaniu trzech kluczowych elementów: kontekstu wprowadzenia miejskiej interwencji (C), mechanizmów wdrożeniowych (M) oraz ich wpływu na efekty tejże interwencji (O). W rzeczywistości termin „teoria programu” zawiera w sobie trzy rodzaje teorii: teorię bazową, teorię zmiany i teorię wdrażania. Ich wzajemne powiązanie przedstawia schemat 3.

TABELA 3. Słabości polityki rozwojowej państwa wraz ze sposobami na ich ograniczenie

Syndrom	Cechy	Konsekwencje	Postulaty
Pasywne przywództwo polityczne	Słabość przywództwa politycznego	Rezygnacja ze strategicznego zarządzania sprawami publicznymi na rzecz biurokratycznego administrowania	<ul style="list-style-type: none"> Wprowadzenie mechanizmów rozliczalności rządzących wykraczających poza procedury wyborcze Budowanie wśród elity rządzącej przekonania, iż rolą państwa w polityce rozwoju jest obniżanie kosztów transakcyjnych procesów rozwojowych, a nie ich szczegółowe projektowanie i operacyjne zarządzanie nimi Wzmocnienie sfery publicznej jako miejsca debaty o sprawach zasadniczych
	Słabość mechanizmów ustanawiania strategicznych celów polityki rozwoju		
	Słabość zaplecza analitycznego i konceptualnego dla prowadzenia polityki rozwoju		
	Rachityczna sfera publiczna i niemrawy dyskurs o sprawach zasadniczych		
Proceduralno-ekstensywny system polityki rozwoju	Utożsamianie wydatkowania środków finansowych z polityką rozwoju	Substytucja zarządzania strategicznego polityką rozwoju polegającą na administrowaniu procedurami	<ul style="list-style-type: none"> Modernizacja instytucji sektora publicznego dla wzmocnienia ich zdolności do realizacji celów strategicznych państwa Kształtowanie warunków do rozwoju kultury administracyjnej zorientowanej na osiąganie celów rozwojowych Budowanie po stronie administracji przekonania o jej słuszności wobec obywateli, organizacji i podmiotów gospodarczych
	Redukowanie kompleksowości przez wzmocnienie proceduralnej złożoności		
	Reifikacja procedur i kult asekuracji		
	Dominacja rozumu urzędniczego nad rozsądkiem		
Niski poziom uczenia się i refleksyjności	Rozkwit komitetów w systemie polityki rozwoju	Niekreatywna adaptacja systemu zarządzania polityką rozwoju	<ul style="list-style-type: none"> Ugruntowanie wśród polityków świadomości znaczenia wiedzy dla dobrego rządzenia Tworzenie warunków dla zmiany kultury organizacyjnej poprzez wprowadzanie zarządczych i technologicznych mechanizmów do prowadzenia polityki opartej na dowodach Intensyfikacja współpracy administracji publicznej z uniwersytetami i ośrodkami badawczymi
	Słabości w sferze analizy, programowania i implementacji oraz oceny polityk publicznych		
	Niski stopień powiązania wyników ewaluacji z procesami decyzyjnymi		
	Dominacja aspektów kontrolno-nadzorczych nad aspektami związanymi z uczeniem się		
	Nadmiernie sformalizowana procedura zlecania badań ewaluacyjnych (badań społecznych)		
	Niski potencjał uczenia się		

Źródło: opracowanie na podstawie Geodecki i in. 2012: 51 i 102.



SCHEMAT 3. Interwencje publiczne jako zestaw teorii – ciąg logiczny

Źródło: Olejniczak 2012: 240.

W omawianej koncepcji teoretycznej identyfikuje się przede wszystkim zależności przyczynowe, których specyfikę trafnie opisuje Joseph A. Maxwell (2004: 245):

Niezależnie od tego, czy przyczynowość „naprawdę istnieje”, czy też nie, to z pewnością istnieje w naszym „świecie codziennego życia” – świecie, w którym realnie żyjemy. Tym, co czyni ją realną w fenomenologicznym sensie, jest możliwość postawienia pytania: „Czy to rzeczywiście jest przyczyna?”. Możliwe jest zweryfikowanie tez o domniemanej przyczynowości, a wraz z tym procesem wytworzenia nowych danych i efektów (...) świat codziennego życia jest pełen przyczyn i efektów.

Innymi słowy, „teoria programu” jest „zbiorem założeń jego twórców na temat związków przyczynowych między podejmowanymi działaniami a pożądanymi efektami, którym towarzyszy na ogół zbiór przekonań dotyczących najbardziej właściwych sposobów wdrożenia interwencji: struktur wdrożenia, sposobu zarządzania, organizacji monitoringu i kontroli finansowej itp.” (Górniak i Mazur 2012: 190).

W udzieleniu odpowiedzi na pytania wyjaśniające wielkość efektu miejskiej interwencji pomocna jest metodologia badań jakościowych, która ma „umożliwić zrozumienie przyczyn sukcesów i porażek leżących po stronie procesów wdrożenia oraz uwarunkowań instytucjonalnych, społecznych i kulturowych” (Górniak i Mazur 2011: 12), a także wskazać, „na jaki rodzaj interwencji można przenieść

doniesienia ewaluacji i co określa stopień przełożenia tych doniesień” (European Cohesion Fund i European Regional Development Fund 2011: 6). Mając dostęp do raportów z ewaluacji wpływu interwencji na poziomie I, można przystąpić do procesu agregacji danych zawartych w tych raportach, czyli przejść na wspomniany wyżej II poziom ewaluacji. Syntezą powinny zostać objęte przede wszystkim dane wyjaśniające zaobserwowane na poziomie I wartości efektów tej samej bądź podobnej interwencji, czyli:

- Wyniki ilościowe rozumiane jako bezpośrednie konsekwencje zastosowanych metod i technik badawczych o charakterze ilościowym, mających na celu wyliczenie wartości efektów netto interwencji (zamierzonych i niezamierzonych).
- Wyniki jakościowe rozumiane jako bezpośrednie konsekwencje zastosowanych metod i technik badawczych o charakterze jakościowym, czyli danych pozyskanych z przeglądu literatury (np. dokumentów programowych, sprawozdawczości czy doniesień prasowych) oraz rozmów z interesariuszami ewaluacji na temat wdrożonej interwencji.
- Wnioski rozumiane jako sformułowanie odpowiedzi na zadane pytania badawcze. W badaniach ewaluacyjnych pytania badawcze mają charakter oceny i z tego powodu badacz wyciągając wnioski, odwołuje się do przyjętych kryteriów ewaluacyjnych. Wnioski (oparte na interpretacji wartości efektów netto) są rozstrzygające i świadczą o sukcesie lub porażce interwencji, a ich omówienie powinno dostarczać informacji wyjaśniających przyczyny zidentyfikowanych wartości efektów interwencji (zamierzonych i niezamierzonych).
- Rekomendacje, czyli praktyczne wskazówki na temat tego, co należy zrobić, aby zwiększyć prawdopodobieństwo sukcesu porównywalnych interwencji w przyszłości. W procesie formułowania konkretnych propozycji zmian badacz może polegać na własnej wiedzy eksperckiej, korzystać z konsultacji ze zleceniodawcą badania i jego interesariuszami, a także z pomocy ekspertów zewnętrznych.

Jednak wykorzystanie wiedzy już wytworzonej (danych zastanych) do procesu budowania bazy wiarygodnych dowodów warunkowane jest przez jakość tej wiedzy.

W polskich realiach identyfikacja czynników kontekstowych i mechanizmów wdrożenia interwencji jest często podejmowana w raportach z ewaluacji *ex-post* ze względu na popularny, operacyjny charakter tego typu badań. Jak argumentuje Karol Olejniczak (2009: 49):

po części taki stan rzeczy był wynikiem wymogów narzuconych Polsce przez Unię Europejską, która oczekiwała, by wykonywane ewaluacje miały charakter dwojakiego

rodzaju: operacyjny i strategiczny. W takiej sytuacji – ze względu na niemożność wypełnienia obu funkcji jednocześnie (...) firmy konsultingowe koncentrowały [większą] uwagę na pomocy w usprawnieniu procesów wdrażania (funkcji operacyjnej) niż na trudniejszej i mniej politycznie delikatnej ocenie efektów programu.

Niestety, nawet w sytuacji, kiedy doniesienia badawcze pozwolą zidentyfikować czynniki relatywnie wpływające na osiągnięte wartości efektów netto interwencji, kluczową kwestią jest wiarygodne uzasadnienie, że faktycznie mamy do czynienia z relacją przyczynową. Metodologiczna wiarygodność doniesień (wyników, wniosków i rekomendacji) ewaluacji zrealizowanych na poziomie I ma krytyczne znaczenie dla zrozumienia, dlaczego miejskie interwencje osiągają planowane i nieplanowane efekty.

Potencjał oceny wiarygodności doniesień zawartych w raportach z ewaluacji miejskich polityk publicznych ma schemat badawczy o nazwie metaewaluacja. Metaewaluacja ocenia wartość i przydatność ustaleń badania ewaluacyjnego, a jej cel polega na ciągłym doskonaleniu praktyk ewaluacyjnych. Koncepcja badań metaewaluacyjnych powstała na gruncie badań edukacyjnych. Po raz pierwszy termin „metaewaluacja” został użyty przez M. Scrivena w 1969 roku raportującego wyniki swojej pracy, mającej na celu ewaluację planu ewaluacji produktów edukacyjnych. Według Scrivena (1991) metaewaluacja oznacza **ewaluację ewaluacji – a pośrednio ewaluację ewaluatorów**, a jej przeprowadzenie powinno być etycznym oraz naukowym obowiązkiem, szczególnie gdy ustalenia ewaluacji mogą wpłynąć na dobrobyt innych osób. Taka sytuacja ma miejsce zwłaszcza w programach o charakterze społecznym. Udostępnianie niedokładnych raportów ewaluacyjnych i/lub nieuprawnionych (obciążonych błędem) ustaleń z oceny systemu polityk publicznych może poważnie wprowadzić w błąd decydentów i skutkować podejmowaniem niewłaściwych decyzji publicznych.

Doprecyzowując, ewaluacja ewaluacji (*evaluating evaluations*) służy do oceny jakości, wiarygodności uzyskanych efektów (Scriven 1991; Stufflebeam 2001; Cooksy i Caracelli 2005), a także wykorzystania do rekomendacji pojawiających się w analizowanych ewaluacjach (Widmer 1996). Przedmiotem analizy są jednak nie odbiorcy wsparcia, ale pojedyncze ewaluacje, które oceniały wypracowane efekty danej interwencji. Badanie ewaluacyjne traktowane jest w kategorii programu/projektu i podlega ewaluacji.

Specyficzne¹⁷ rozumienie metaewaluacji wyrosło na gruncie funduszy strukturalnych i ogranicza metaewaluację do syntezy wniosków i rekomendacji z ewaluacji (Bienias 2011). Jak argumentują Karol Olejniczak i Marek W. Kozak (2011: 55):

¹⁷ Specyficzne, bo odbiegające od tradycyjnego rozumienia badania metaewaluacyjnego w edukacji na świecie.

Na poziomie krajowym coraz wyraźniejsza jest potrzeba metaewaluacji – a więc badań, które porównują, zestawiają i dokonują syntezy wniosków z wielu podobnych badań tematycznych. Aby jednak móc prowadzić metaewaluację, konieczna jest wysoka dyscyplina w pisaniu i wnioskowaniu na poziomie pojedynczych badań. Pojedyncze badania programów muszą spełniać pewne standardy metodyczne.

Ograniczenie zastosowania „metaewaluacji” tylko do syntezy doniesień będzie możliwe wówczas, kiedy powszechną praktyką będzie przeprowadzanie badań ewaluacyjnych wysokiej jakości, opartych na profesjonalnych, ogólnodostępnych standardach i wymogach metodologicznych. W takiej sytuacji realizacja głównego celu metaewaluacji, a więc oceny jakości doniesień ewaluacji, będzie mogła zostać ograniczona do niezbędnego minimum.

Obecnie nie ma mowy o takiej praktyce. Henryk Mizerek (2002) twierdzi wręcz, że:

praktyki upowszechniane w Polsce bardzo często pomijają fazę projektowania. Początkiem większości ewaluacji jest poszukiwanie gotowych narzędzi badawczych. W rezultacie mamy do czynienia z ankietomanią na masową skalę. Ankietuje się wszystkich i wszędzie. Nierzadko ewaluacja zaczyna i kończy się na tym etapie. Brakuje czasu, umiejętności i możliwości, by zebraną w ten sposób wielką ilość informacji poddać analizie i przedstawić w formie raportu. Nawiasem mówiąc, może to i dobrze – albowiem są to często informacje zupełnie bezużyteczne.

Z tego powodu osoby zarządzające miejskimi programami publicznymi powinny koncentrować się przede wszystkim na zbudowaniu zaplecza badawczego do klasycznie rozumianego badania metaewaluacyjnego, czyli **sposobów na weryfikację jakości i wiarygodności ustaleń badań ewaluacyjnych (główny cel metaewaluacji)**. Takie działania zapewnią podstawę syntezy i porównywania wyników, wniosków oraz rekomendacji zawartych w badaniach oceniających efekty interwencji. Uważam, że dobra ewaluacja nie powinna bronić się przed ewaluacją samej siebie. Tylko w ten sposób zrobimy krok do przodu w dostarczaniu naukowych dowodów na potrzeby strategicznego i lepszego zarządzania miastem.

Bibliografia

- Babbie Earl, 2003, *Badanie społeczne w praktyce*, przeł. Witold Betkiewicz, Marta Bucholc, Przemysław Gadomski, Jacek Haman, Agnieszka Jasiewicz-Betkiewicz, Agnieszka Kłoskowska-Dudzińska, Michał Kowalski i Maja Mozga, Warszawa: PWN.
- Bennett Gavin i Nasreen Jessani (red.), 2011, *The Knowledge Translation Toolkit*, New Delhi: Sage India.

- Bienias Stanisław, 2011, *Metaewaluacja jako narzędzie służące projektowaniu przyszłych interwencji publicznych. Przykład zastosowania w ramach Polityki Spójności*, [w:] Agnieszka Haber i Rafał Trzciński (red.), *Ocena wpływu i prognozowanie efektów w badaniach ewaluacyjnych*, Warszawa: Polska Agencja Rozwoju Przedsiębiorczości, s. 9–22.
- Brouwer E.B., P. den Hertog, T. Poot i J. Segers, 2002, *WBSO nader beschouwd; onderzoek naar de effectiviteit van de WBSO*, Amsterdam: PriceWaterhouseCoopers, Dialogic, TU Delft.
- Carr David, 1992, *Practical Enquiry, Values and the Problem of Educational Theory*, „Oxford Review of Education” t. 18, nr 3, s. 241–251.
- Cooksy Leslie J. i Valerie J. Caracelli, 2005, *Quality, context & use: issues in achieving the goals of metaevaluation*, „American Journal of Evaluation” t. 26, s. 31–42.
- Davies Philip, 2006, *What is needed from research synthesis from a policy-making perspective?*, [w:] Jennie Popay (red.), *Moving beyond effectiveness in evidence synthesis: Methodological issues in the synthesis of diverse sources of evidence*, London: National Institute for Health and Clinical Excellence, s. 97–104.
- DG Budget Evaluation Unit, 2004, *Evaluating EU activities. A practical guide for the commission services*, European Commission.
- Dyrekcja Generalna XIX – Budżet, 1997, *Ewaluacja programów wydatków Unii Europejskiej. Przewodnik. Ewaluacja w połowie okresu i ex-post*, przeł. Urząd Komitetu Integracji Europejskiej i francuskie Krajowe Centrum Terytorialnej Służby Cywilnej (nieoficjalne tłum.), Komisja Europejska.
- European Cohesion Fund i European Regional Development Fund, 2011, *The programming period 2014–2020. Guidance document on monitoring and evaluation. Concepts and recommendations*, Bruxelles: European Commission.
- European Commission, 2006, *Indicative Guidelines on Evaluation Methods: Ex ante Evaluation*, Working Document nr 1.
- Geodecki Tomasz, Grzegorz Gorzelak, Jarosław Górniak, Jerzy Hausner, Stanisław Mazur, Jacek Szlachta i Janusz Zalewski, 2012, *Kurs na innowacje. Jak wyprowadzić Polskę z rozwojowego dryfu?*, Kraków: Fundacja Gospodarki i Administracji Publicznej.
- Górniak Jarosław, Barbara Worek, Weronika Orkisz, Zuzanna Drożdżak i Patrycja Antosz, 2010, *Ewaluacja oparta na badaniu użyteczności – analiza w kontekście realizacji zasady good governance w wybranych programach operacyjnych*, Kraków: Centrum Ewaluacji i Analiz Polityk Publicznych.
- Górniak Jarosław, 2009, *Efekt przyczynowy w ewaluacji*, [w:] Agnieszka Haber i Maciej Szałaj (red.), *Ewaluacja wobec wyzwań stojących przed sektorem finansów publicznych*, Warszawa: Polska Agencja Rozwoju Przedsiębiorczości, s. 192–197.
- Górniak Jarosław i Karolina Keler, 2008, *Ewaluacja i wskaźniki w cyklu planowania i realizacji polityk publicznych*, „Zarządzanie Publiczne” nr 1(3), s. 109–124.
- Górniak Jarosław i Stanisław Mazur, 2012, *Analiza polityk publicznych i programowanie w obszarze strategii rozwoju*, [w:] Jarosław Górniak i Stanisław Mazur (red.), *Zarządzanie strategiczne rozwojem*, Warszawa: Ministerstwo Rozwoju Regionalnego, s. 185–220.
- Górniak Jarosław i Stanisław Mazur, 2011, *Polityki publiczne oparte na dowodach i ich zastosowanie do rynku pracy*, [w:] Jarosław Górniak i Stanisław Mazur (red.), *W kierunku polityki rynku pracy opartej na dowodach*, Warszawa: Obserwatorium Regionalnych Rynków Pracy Pracodawców Rzeczypospolitej Polskiej, s. 9–34.
- Heckman James J. i Jeffrey A. Smith, 1996, *Experimental and Nonexperimental Evaluation*, [w:] Günther Schmid, Jacqueline O'Reilly i Klaus Schömann (red.), *International Handbook of Labour Market Policy and Evaluation*, Cheltenham: Edward Elgar, s. 37–88.

- Lomas Jonathan, Tony Culver, Chris McCutcheon, Laura McAuley i Susan Law, 2005, *Conceptualizing and Combining evidence for health system guidance*, Ottawa: Canadian Health Services Research Foundation.
- Maxwell Joseph A., 2004, *Using Qualitative Methods for Casual Explanation*, „Field Methods” t. 16 (3), s. 243–264.
- Ministerstwo Rozwoju Regionalnego, 2007, *Narodowe Strategiczne Ramy Odniesienia na lata 2007–2013. Wytyczne nr 6 w zakresie ewaluacji programów operacyjnych na lata 2007–2013*, 200MRR/H/6(1)05/2007, Warszawa.
- Mizerek Henryk, 2002, *Ewaluacja w edukacji – szansa z (kilkoma) pułapkami w tle. Nauczyciel i uczeń w przestrzeniach szkoły*, [w:] Marzena Nowicka (red.), *Nauczyciel i uczeń w przestrzeniach szkoły*, Olsztyn: Wydawnictwo Uniwersytetu Warmińsko-Mazurskiego.
- Mosselman Marco i Yvonne Prince, 2004, *Review of Methods to Measure the Effectiveness of State Aid to SME's. Final report to the European Commission*, Zoetermeer: EIM.
- OECD–DAC, 2002, *Glossary of Key Terms in Evaluation and Results Based Management*, Paris: OECD Publications.
- Olejniczak Karol, 2009, *Praktyka ewaluacji efektów programów rozwoju regionalnego – studium porównawcze*, Raporty i analizy EUROREG 2.
- Olejniczak Karol, 2012, *Monitoring i ewaluacja*, [w:] Jarosław Górniak i Stanisław Mazur (red.), *Zarządzanie strategiczne rozwojem*, Warszawa: Ministerstwo Rozwoju Regionalnego, s. 239–259.
- Olejniczak Karol i Marek W. Kozak, 2011, *Do czego warto wykorzystać ewaluację? Pożądane kierunki ewaluacji polityki rozwoju*, [w:] praca zbiorowa, *Badania ewaluacyjne – po co? Budowa przyszłości regionu*, Opole: Urząd Marszałkowski Województwa Opolskiego, Departament Polityki Regionalnej i Przestrzennej, Jednostka Ewaluacyjna, s. 41–58.
- Pincus Allen i Anne Minahan, 1973, *Social Work Practice: Model and Method*, Itasca: F.E. Peacock Publishers.
- Portal Funduszy Europejskich, http://www.ewaluacja.gov.pl/slownik/Strony/slownik_efekt_przeniesienia.aspx [dostęp: 20.07.2012].
- Scriven Michael, 1972, *Pros and cons about goal free evaluation*, „Evaluation Comment” 3(4), 1–7.
- Scriven Michael, 1991, *Evaluation thesaurus* (4th ed.), Newbury Park: Sage.
- Słownik języka polskiego PWN, www.sjp.pwn.pl [dostęp: 20.09.2017].
- Storey David, 2002, *Six Steps to Heaven: Evaluating the Impact of Public Policies to Support Small Business in Developed Economies*, „International Journal of Entrepreneurship Education” t. 1, nr 2, s. 181–202.
- Stufflebeam Daniel L., 2001, *The Metaevaluation Imperative*, „American Journal of Evaluation” t. 22(2), s. 183–209.
- Tyler Ralph W., 1942, *General statement on evaluation*, „The Journal of Educational Research” t. 35, nr 7, s. 492–501.
- Widmer Thomas, 1996, *Meta-Evaluation. Kriterien zur Bewertung von Evaluationen*, Stuttgart–Wien: Verlag Paul Haupt Bern.